What is claimed is:

1.   A videoconferencing device comprising:

a video sensor for capturing an image;

a plurality of microphones for generating an audio signal in response to an acoustic source; and

a processing engine coupled to the video sensor and the plurality of microphones for generating at least one video stream and a position signal indicating a position of the acoustic source.

2.   The videoconferencing device of claim 1, further comprising a phase synchronization engine coupled to the video sensor for synchronizing a phase between the video sensor and a video display output.

3.   The videoconferencing device of claim 1, further comprising a communication interface coupled to the processing engine for transmitting the audio signal, position signal, and at least one video stream to a remote videoconferencing device.

4.   The videoconferencing device of claim 1, wherein the position signal is generated based upon magnitude differences of electric or current signals received from the plurality of microphones.

5.   The videoconferencing system of claim 1, wherein the processing engine further comprises a video processing engine, the video processing engine defining a plurality of image sections and generating a respective plurality of video streams corresponding to the plurality of image sections.

6.   The videoconferencing system of claim 1, wherein if the position of the sound source changes, the processing engine generates a new position signal to reflect a position change.

7.   The videoconferencing device of claim 2, wherein the remote videoconferencing device selectively drives one or more speakers in response to the position signal to play the audio signal corresponding to the image of the at least one video stream.

8.   The videoconferencing device of claim 1, wherein the plurality of microphones are arranged in an n-fire configuration.

9.   The videoconferencing device of claim 1, wherein the plurality of microphones are arranged in a vertical array.

10.   The videoconferencing device of claim 5, wherein the processing engine scale a first image section of the plurality of image sections into a first video stream having a first resolution and scales a second image section of the plurality of image sections into a second video stream having a second resolution.

11.   The videoconferencing system of claim 1, further comprising a pan motor coupled to the video sensor for providing a larger degree view angle.

12.   A method for transmitting conferencing data in a video conferencing system, comprising:

capturing an image with a video sensor and generating at least one video stream from the image;

capturing audio data with a plurality of microphones and generating an audio signal;

generating a position signal indicating a position of an acoustic source based upon magnitude differences of the audio data; and

transmitting the position signal, audio signals, and the at least one video streams via a communication channel.

13.   The method of claim 12, further comprising selectively driving one or more speakers of a remote video conferencing system in response to the position signal to play the audio signal corresponding to the image of the at least one video stream.

14.   The method of claim 12, further comprising synchronizing a phase between the video sensor and a video display output.

15.   The method of claim 12, further comprising defining a plurality of image sections and generating a respective plurality of video streams corresponding to the plurality of image sections.

16.   The method of claim 12, further comprising generating a new position signal to reflect a position change.

17.   The method of claim 14, further comprising scaling a first image section of the plurality of image sections into a first video stream having a first resolution and scaling a second image section of the plurality of image sections into a second video stream having a second resolution.

18.   A videoconferencing device comprising:

means for capturing an image and generating at least one video stream from the image;

means for capturing audio and generating an audio signal;

means for generating a position signal indicating a position of an acoustic source based

upon magnitude differences of the audio data, the position signal selectively

driving one or more speakers of a remote videoconferencing system in response to

the position signal to play the audio signal corresponding to the image of the at

least one video stream; and

means for transmitting the position signal, audio signals, and the at least one video

streams via a communication channel.

19.   An electronically-readable medium having embodied thereon a program, the program

being executable by a machine to perform method steps for transmitting conferencing data, the

method steps comprising:

capturing an image with a video sensor and generating at least one video stream from the

image;

capturing audio data with a plurality of microphones and generating an audio signal;

generating a position signal indicating a position of an acoustic source based upon

magnitude differences of the audio data; and

transmitting the position signal, audio signals, and the at least one video streams via a

communication channel.

20.   The electronically-readable medium of claim 19, wherein the method steps further

comprise selectively driving one or more speakers of a remote videoconferencing system in

response to the position signal to play the audio signal corresponding to the image of the at least

one video stream.

21.   The electronically-readable medium of claim 19, wherein the method steps further

comprise defining a plurality of image sections and generating a respective plurality of video

streams corresponding to the plurality of image sections.

22.   The electronically-readable medium of claim 19, wherein the method steps further comprise scaling a first image section of the plurality of image sections into a first video stream having a first resolution and scaling a second image section of the plurality of image sections into a second video stream having a second resolution.